

---

# Nested Neural Networks

---

Yoram Baram

---

(NASA-TM-101032) NESTED NEURAL NETWORKS  
(NASA) 26 p CSCI 09B

N88-30373

Unclas  
G3/63 0159385

August 1988

---

# **Nested Neural Networks**

---

Yoram Baram, Ames Research Center, Moffett Field, California

August 1988



National Aeronautics and  
Space Administration

**Ames Research Center**  
Moffett Field, California 94035

# Nested Neural Networks

Yoram Baram\*

## Abstract

Nested neural networks, consisting of small interconnected subnetworks, allow for the storage and retrieval of neural state patterns of different sizes. The subnetworks are naturally categorized by layers corresponding to spatial frequencies in the pattern field. The storage capacity and the error correction capability of the subnetworks generally increase with the degree of connectivity between layers (the “nesting degree”). Storage of only few subpatterns in each subnetwork results in a vast storage capacity of patterns and subpatterns in the nested network, maintaining high stability and error correction capability.

---

\*Senior Research Associate of the National Research Council at the NASA Ames Research Center, Moffett Field, CA 94035, on sabbatical leave from the Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel.

# 1 Introduction

Biological neural networks are generally perceived as large populations of interconnected brain cells. Their anatomy and physiology are only partly known or understood, yet certain empirically observed characteristics seem widely recognized (see a recent account by Crick and Asanuma [1]). The neuron's state, its membrane potential, depends on the states of other neurons connected to it and on an external potential induced through some interface by a sensory device. Inter-neural connections consist of fibers transmitting electric charges from one neuron to another through synaptic contacts. Not all neurons are directly interconnected, yet every two neurons are connected by a small number of indirect connections. Most of the inter-neuron fibers extend for relatively short distances (less than 1 mm), but a significant number spread for longer distances. A typical cortical area is not connected directly to all or even most other areas. Information processing, as evidenced by neural firing activity, is distributed throughout the cortex. Single neurons or groups of neurons may be eliminated without causing a noticeable degradation in performance.

Neural information processing models largely assume that information is stored in the synaptic contacts and conveyed by patterns of neural states (see, e.g., Rosenblatt [2], Kohonen [3] and Grossberg [4]). A model based on the Hebbian storage mechanism ([5], see also Cooper [6]) and the McCulloch-Pitts retrieval mechanism [7] was shown by Hopfield [8] to possess a certain associative memory capability. The model consists of  $N$  interconnected neurons, each capable of being in one of two states, represented by the values  $\pm 1$ . Denoting the state of the  $i$ 'th neuron by  $x_i$ , the state of the network is defined as the vector  $x = (x_1, \dots, x_N)^T$ , where  $(\cdot)^T$  denotes transpose. The information in  $M$  given patterns,  $x^{(1)}, \dots, x^{(M)}$ , is stored in synaptic parameters, calculated according to the Hebbian learning rule as

$$T_{i,j} = \sum_{l=1}^M x_i^{(l)} x_j^{(l)} \quad \text{or} \quad T = \sum_{l=1}^M x^{(l)} x^{(l)T} \quad (1)$$

However, in the Hopfield model  $T_{i,i} = 0$ ,  $i = 1, \dots, N$ . Information is retrieved when the network is probed by some pattern, which constitutes the initial state. The neural states are then updated, one at a time, according to the McCulloch-Pitts rule

$$x_i(n+1) = \begin{cases} +1 & \text{if } \sum_{j=1}^N T_{i,j} x_j(n) \geq 0 \\ -1 & \text{if } \sum_{j=1}^N T_{i,j} x_j(n) < 0 \end{cases} \quad (2)$$

While the neuron state values were assumed to be binary [8], a biologically more accurate graded response model was shown to behave in essentially the same manner as the binary

one [9]. Each neuron was generally assumed to be connected to all the other neurons by two-directional connections. The possibility of disconnected neurons was addressed by simply substituting zeros for the corresponding synaptic parameter value. However, no specific partial connectivity patterns or potential advantages thereof have been mentioned.

The network was shown by Hopfield [8] to be globally stable in the sense that, when probed by any pattern, it converges to a final state, which locally minimizes the energy function  $-x^T T x$ . Such stability was shown to follow from the symmetry of  $T$  and the asynchronous neural update rule (2). Hopfield also found empirically that a network of 100 neurons converges to the stored pattern nearest to the probe without severe error if it stores less than 15 patterns and every two patterns differ by  $50 \pm 5$  bits [8]. The latter requirement is equivalent to near-orthogonality in the Euclidean sense, which does not seem to be readily resolvable for network-size patterns. The storage capacity of the network has been defined as the number of stored patterns that are equilibrium points (see, e.g., Abu-Mostafa and St.Jacques [10]). McEliece et al. [11] and Bruce et al. [12] independently showed, using large sample assumptions, that for random stored patterns to be equilibrium points of the network with finite probability, their number cannot exceed  $N/(2\log N)$ , where  $N$  is the number of neurons. Montgomery and Vijayakumar [13] showed that the maximal number of stored patterns that can be guaranteed to be equilibrium points of the network for the worst choice of such patterns is 1 for zero-valued diagonal synaptic parameters, independently of the network size. It was further shown by Dembo [14] that the worst case capacity for nonzero diagonal parameters is 2, for  $N \geq 6$  (It is shown in the present paper that the maximal capacity for the latter case is 2 for  $N = 2$ , 3 for  $N = 3$  and 2 for  $N \geq 4$ ). It is known that if the stored patterns are restricted to be mutually orthogonal, they are equilibrium points of the network. The number of such patterns is, of course, bounded by  $N$  and the network loses its error correction capability as the number of orthogonal stored patterns reaches  $N$  (see, e.g. Horn and Weyers [15]). A storage rule that guarantees that the stored patterns are equilibrium points of the network will be called a "safe storage" rule.

The network structures considered in this work were largely motivated by the observation that while the safe storage of network-size patterns in a large fully connected neural network is either unresolvable (if only orthogonal patterns are to be stored) or of little practical use (if only two patterns are to be stored), the safe storage of subpatterns in relatively small subnetworks can be sensibly accomplished and will result in a vast storage capacity of the entire network. The sizes of the subpatterns are naturally categorized by spatial periods or frequencies in the pattern field. The corresponding subnetworks are

subsequently categorized by layers, each corresponding to a spatial frequency. Inter-layer connections are suggested by the fact that points in the pattern field may be shared by subpatterns of different sizes. Connecting each neuron at a given layer to its nearest neighbors in lower layers results in the nesting of subnetworks corresponding to neighboring layers and sharing a common neuron. The number of consecutive layers connected in this fashion defines the network's "nesting degree". The emerging network structures resemble the "fractal" forms studied by Mandelbrot [16], who conjectured, in general terms, the fractal geometry of neural networks. He wrote: "The Purkinje cells in mammalian cerebellum are practically flat and their dendrites form a plane-filling maze. From mammals to pigeon, alligator, frog and fish, the degree of filling decreases (Llinas 1969). It would be nice if this corresponded to a decrease in  $D$  [the Hausdorff-Besicovich dimension], but the notion that neurons are fractal remains conjectural". I have chosen to use the term "nested" rather than fractal, not only because the former seems more descriptive and familiar than the latter, but, in particular, because the "nesting degree", as defined in this work, seems central to the proposed network structures. While the specific fractal structures examined by Mandelbrot, such as the Koch tree, do not allow for nesting degrees higher than 2, certain similar structures allow, as shown in this work, for higher nesting degrees.

The storage and retrieval of information in a nested neural network are naturally accomplished in the form of subpatterns. The storage capacity and the error correction capability of the subnetworks in a nested network are shown to generally increase with the nesting degree. Storage of only few patterns in each subnetwork is suggested by both physical and information retrieval considerations. The worst-case capacity of nested subnetworks is the same as that of externally disconnected networks of the same size and their error correction capability is generally higher. The probability of orthogonality between random patterns is shown to be approximately inversely proportional to the square root of the pattern size. It follows that the orthogonality condition can be sensibly met in the storage of subpatterns in a nested network, using a simple threshold function. For random subpatterns stored in relatively large subnetworks, the probabilities of local stability and error correction are shown to increase with the nesting degree. The fact that large patterns often share some or many of their subpatterns implies that the number of different network-size patterns that can be stored and retrieved in the nested network is exponential in the number of the subnetworks. Storage of relatively few subpatterns in each subnetwork results in a vast storage capacity of the entire network, maintaining the stability and error correction capability of the subnetworks. The nested structure also allows for the retrieval of individual subpatterns. It requires considerably less wires and connection devices than fully connected networks, and allows for the local reconstruction

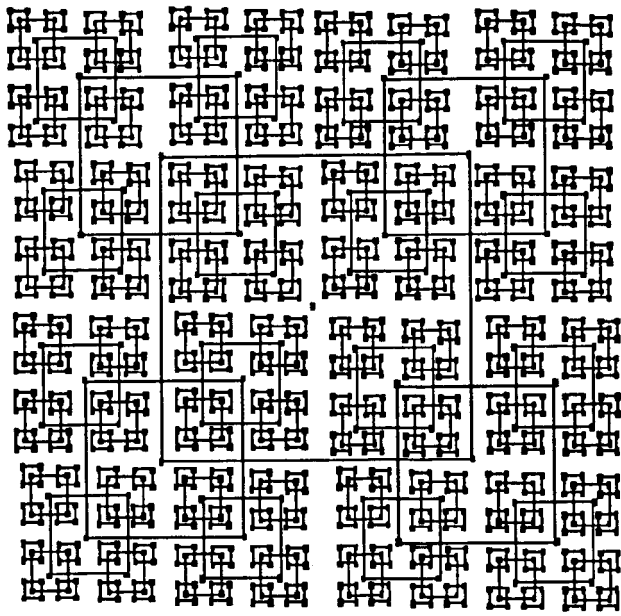
of damaged subnetworks without rewiring the entire network.

## 2 Nested Neural Networks

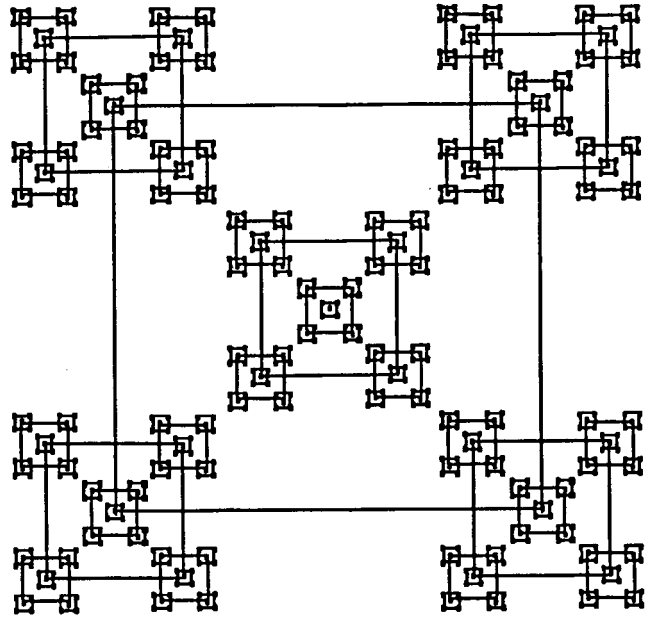
Let a pattern field be defined by the sample points in a bounded regime of a multidimensional Euclidean space, sampled according to some sampling geometry (e.g., rectangular or hexagonal) at  $P$  different spatial frequencies (see, e.g., Dudgeon and Mersereau [17]). Suppose that a neuron is placed at each sample point and, for positive integers  $K_p$ , let the neurons corresponding to the  $p$ 'th frequency be arranged in groups of  $K_p - 1$  fully interconnected neighboring neurons. Suppose that, for some positive integer  $S$ , each neuron at the  $i$ 'th layer is connected to its  $K_p - 1$  neighbors in the  $p$ 'th layer, for each  $p$  in the range  $i - S \leq p \leq i$ , if  $i \geq S$ , or in the range  $p < i$ , if  $i < S$ . Each neuron at the  $i$ 'th layer is then a member of  $S$  subnetworks corresponding to  $S$  consecutive lower layers (including the  $i$ 'th layer), if  $i \geq S$ , or of  $i$  subnetworks corresponding to all the lower layers, if  $i < S$ . In the emerging structure, subnetworks corresponding to lower layers (higher spatial frequencies) are "nested" in subnetworks corresponding to higher layers. The parameter  $S$  may then be characterized as the nesting degree of the network.

Connecting each neuron at a given layer to its nearest neighbors in the neighboring lower layer results in a simple connectivity pattern between layers, with each neuron in every layer except the lowest one being a member of two subnetworks corresponding to neighboring spatial frequencies. Such network structures may be generated by branching, much like the Koch tree ([16]). Figure 1(a) shows such a nested network of 5 layers and 4 or 5 neurons in each subnetwork, depending on whether the layers are connected ( $S = 2$ ) or disconnected ( $S = 1$ ). In this network, neurons are located at the four corners and the center of each square. The smallest squares represent the subnetworks of the first layer, corresponding to the highest spatial frequency. Larger squares represent higher layers of the network, corresponding to lower frequencies. For graphical clarity the figure does not detail all the inter-neural connections within the subnetworks, which may be generally assumed to be fully connected. In contrast to the network in Figure 1(a), which does not allow for nesting degrees higher than 2, the network in Figure 1(b) allows for higher nesting degrees. The latter can be seen to be sparse in the Euclidean topology, and may not be suitable for the direct sampling of, say, a visual pattern field. As shown next, this sparsity can be removed by choosing different sampling geometries.

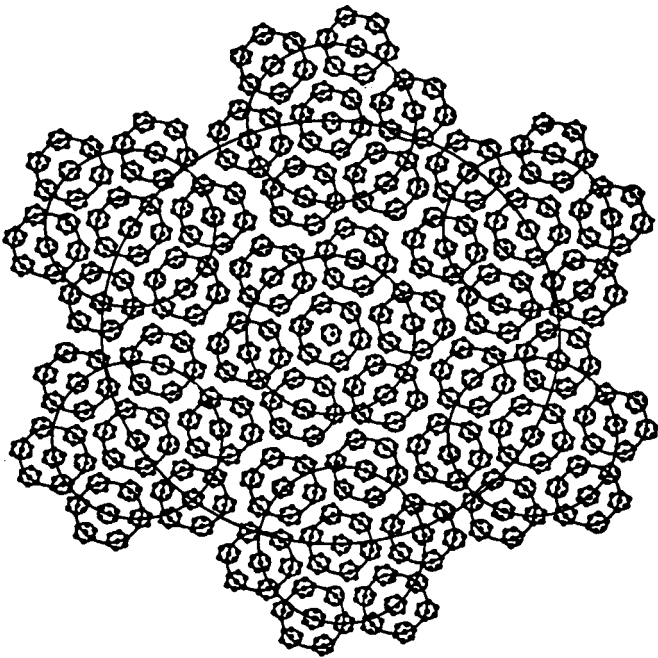
Figure 1(c) shows a nested neural network of hexagonal structure. The hexagonal



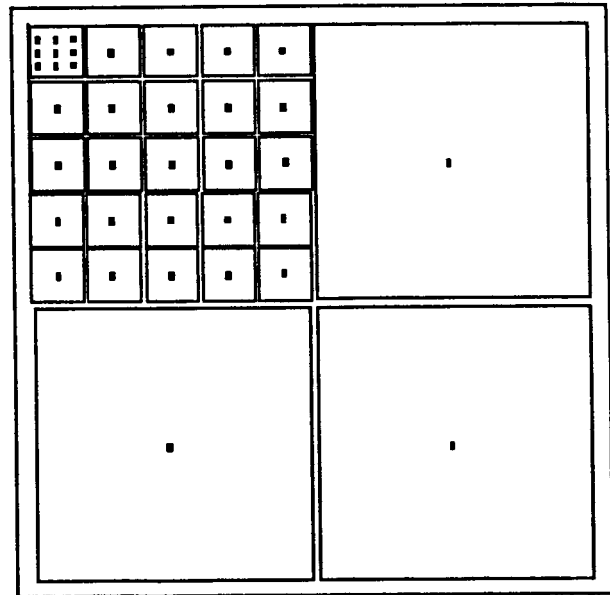
(a)



(b)



(c)



(d)

Figure 1: Nested network structures



sampling geometry is known to be the most efficient from the viewpoints of space filling and multidimensional bandlimited signal representation (see [17]). Hexagonal lattice structures have been observed in the primate retina and visual cortex (see, e.g., Ahumada [18] and Watson and Ahumada [19]). An exact construction of such structures can be obtained using the hexagonal sampling matrix (see, e.g., [17], [19]). It can be seen that in this network each neuron at a given layer can become a member of subnetworks in any number of lower layers when it is connected to its nearest neighbors in those layers. The hexagonal structure also provides a high degree of spatial symmetry, which would be required in order to achieve the property of rotation invariance in visual pattern recognition. While the hexagonal structure is optimal from the sampling viewpoint, nested networks of larger sizes would provide, as shown in the next section, higher error correction capabilities. Such networks can be obtained using rectangular periodic sampling ([17], p. 63), which also allows for the generalization of the nested structure into one with different subnetwork sizes for each layer, maintaining relatively compact packing and, as in the hexagonal structure, high nesting degrees. Figure 1(d) depicts such a network consisting of 3 layers of subnetwork sizes 9, 25 and 4. The subnetworks are represented by squares. For graphical clarity, the neurons of only one subnetwork in each layer are shown, without the connections within the subnetwork, which may be assumed to be fully connected.

The determination of a sensible sampling rule is, then, a key to the design of a nested neural network. Given a sampling geometry, the design requires the determination of the sampling frequencies. These can be determined by first performing spectral estimation on a given set of "training patterns", viewing them as bandlimited multidimensional signals, and then applying the sampling theorem (see, e.g., [17]). The construction rule for a nested network is simple: duplicate (branch out) and connect. The basic unit, a neuron, is duplicated a number of times and the resulting group is connected into a new unit, which is duplicated by the same rule into a group which is connected and duplicated, etc.. The resulting network can withstand extensive damage because of the high degree of duplicity. Moreover, damaged neurons or subnetworks can be repaired or recreated locally without recreating or rewiring the whole network, which stands in contrast to fully connected networks. Also, nested networks require considerably less wires (fibers, "white matter") and connection devices (synapses) than fully connected ones.

### 3 Information Storage and Retrieval

Let the state of the  $k$ 'th neuron be denoted by  $x_k$  and suppose that it can only take the binary values  $\pm 1$ . Denoting the state of the network by  $x$ , let  $x_{(i)}^{(j)}$  define a vector arranged in the same order as  $x$ , whose components corresponding to the  $i$ 'th subnetwork are the same as those of the  $j$ 'th subpattern stored in that subnetwork and the others are zero. Although  $x_{(i)}^{(j)}$  is a network-size vector, the information it contains is essentially concentrated in the few bits corresponding to the  $i$ 'th subnetwork and, subsequently, it will be referred to as a "subpattern". Since the zero components of a subpattern do not affect the corresponding synaptic parameters, the subpattern  $x_{(i)}^{(j)}$  may be viewed as being stored exclusively in the  $i$ 'th subnetwork. The matrix of synaptic parameters for the nested network, corresponding to the Hebbian storage rule, can then be written as

$$T = \sum_{i=1}^L \sum_{j=1}^{M_i} x_{(i)}^{(j)} x_{(i)}^{(j)T} \quad (3)$$

where  $M_i$  denotes the number of subpatterns stored in the  $i$ 'th subnetwork and  $L$  denotes the number of subnetworks. The resulting matrix  $T$  is symmetric and highly sparse. Information retrieval is initiated by a probe (initial state)  $x(0)$ . Neurons are then selected at random, one at a time, and their states are updated according to the McCulloch-Pitts rule (2). The storage mechanism (3) may be replaced by one in which the diagonal elements  $T_{i,i}$  are forced to be zero, as in the Hopfield model [8]. Replacing the threshold 0 in the retrieval rule (2) by the threshold  $-T_{i,i}x_i(n)$  will then result in a network equivalent to one with the storage rule (3).

Global stability, that is, convergence to some final state, is implied by the symmetry of the matrix of synaptic parameters, as in the case of fully connected networks. For a stored subpattern to be retrievable, it is necessary that it is an equilibrium point of the network that, once reached, is never left. Suppose that the  $k$ 'th neuron is to be updated and that this neuron is a member of  $S_k$  nested subnetworks. Further suppose that the  $i$ 'th of these subnetworks stores  $M_i$  subpatterns and that its state is a stored subpattern  $x_{(i)}^{(l_i)}$ . Since the  $k$ 'th neuron is only connected to neurons in the  $S_k$  subnetworks, we have

$$[Tx]_k = \left[ \sum_{i=1}^{S_k} \sum_{j=1}^{M_i} x_{(i)}^{(j)} x_{(i)}^{(j)T} x_{(i)}^{(l_i)} \right]_k \quad (4)$$

The neuron in question obviously has the same value in all the subpatterns  $x_{(i)}^{(l_i)}$ ,  $i =$

$1, \dots, S_k$ , hence,

$$[Tx]_k = S_k K x_k + \left[ \sum_{i=1}^{S_k} \sum_{\substack{j=1 \\ j \neq l_i}}^{M_i} x_{(i)}^{(j)} x_{(i)}^{(j)T} x_{(i)}^{(l_i)} \right]_k \quad (5)$$

where  $K$  is the number of neurons in each of the subnetworks, assumed for convenience to be uniform throughout the network. The  $k$ 'th neuron will not change if and only if the sign of  $[Tx]_k$  is the same as that of  $x_k$ . This will be the case if and only if the second term of (5) does not offset the first. The stored subpatterns will be equilibrium points of the network if and only if this is true for each of the neurons. If a subnetwork stores replicas of a stored subpattern  $x_{(i)}^{(l_i)}$  or its converse  $-x_{(i)}^{(l_i)}$ , the corresponding terms will be added to the first term in (5), increasing it with respect to the second. This implies that if a stored subpattern is an equilibrium point of the network when the corresponding subnetwork does not store its replicas or converses, it will also be an equilibrium point when any number of its replicas and converses are stored in the same subnetwork.

Denoting by  $d[x_{(i)}^{(j)}, x_{(i)}^{(l)}]$  the Hamming distance between  $x_{(i)}^{(j)}$  and  $x_{(i)}^{(l)}$  and by  $r[x_{(i)}^{(j)}, x_{(i)}^{(l)}] = d[x_{(i)}^{(j)}, x_{(i)}^{(l)}]/K$  the proportional distance, (5) can be written as

$$[Tx]_k = S_k K x_k + \sum_{i=1}^{S_k} \sum_{\substack{j=1 \\ j \neq l_i}}^{M_i} K(1 - 2r[x_{(i)}^{(j)}, x_{(i)}^{(l_i)}])[x_{(i)}^{(j)}]_k \quad (6)$$

an offset of the first term of (6) will not occur even in the worst case if and only if

$$\sum_{i=1}^{S_k} \sum_{\substack{j=1 \\ j \neq l_i}}^{M_i} |(1 - 2r[x_{(i)}^{(j)}, x_{(i)}^{(l_i)}])| < S_k \quad (7)$$

This is the necessary and sufficient condition for the stored subpatterns to be equilibrium points of the network. When the layers are disconnected (nesting degree 1), the necessary and sufficient condition is that for each of the subnetworks

$$\sum_{\substack{j=1 \\ j \neq l_i}}^{M_i} |(1 - 2r[x_{(i)}^{(j)}, x_{(i)}^{(l_i)}])| < 1$$

The latter is a sufficient but not necessary condition for (7). This implies that inter-layer connections potentially increase the likelihood that the stored subpatterns are equilibrium points of the network.

Suppose that the  $i$ 'th subnetwork stores  $M_i$  subpatterns and that an additional subpattern  $x_{(i)}^{(M_i+1)}$  is to be stored. Let us denote by  $T(M_i)$  and  $T(M_i + 1)$  the synaptic

parameter matrix before and after the new storage, respectively. Also let  $I_i$  denote the index set of the neurons of the  $i$ 'th subnetwork. Then, for each  $k \in I_i$ , we have

$$[T(M_i + 1)x_{(i)}^{(M_i+1)}]_k = S_k K [x_{(i)}^{(M_i+1)}]_k + [T(M_i)x_{(i)}^{(M_i+1)}]_k$$

For the last subpattern to be an equilibrium point of the network, it is necessary that

$$|[T(M_i)x_{(i)}^{(M_i+1)}]_k| < S_k K \quad (8)$$

The latter represents a storage constraint, which redefines the storage rule to be

$$T(M_i + 1) = \begin{cases} T(M_i) + x_{(i)}^{(M_i+1)} x_{(i)}^{(M_i+1)T} & \text{if } |[T(M_i)x_{(i)}^{(M_i+1)}]_k| < S_k K, \quad k \in I_i \\ T(M_i) & \text{otherwise} \end{cases} \quad (9)$$

It should be noted that condition (8) is necessary even when additional subpatterns are stored in the  $i$ 'th subnetwork after the  $M_i + 1$ 'st subpattern, if the latter is to be an equilibrium point even in the worst case, as these additional subpatterns may be orthogonal to it. The condition will be satisfied under the safe storage mechanisms discussed in the sequel and may be imposed on the storage of random subpatterns, increasing the probability that these subpatterns become equilibrium points of the network.

Even when the stored subpatterns are equilibrium points of the network, they do not necessarily attract neighboring states. Such local attraction may be characterized as an error correction capability, as it will provide convergence to the correct subpattern, when the corresponding subpattern in the probe represents a sufficiently small diversion from a stored subpattern. Suppose that the  $k$ 'th neuron is to be updated and that it is a member of  $S_k$  nested subnetworks, the  $i$ 'th of which stores  $M_i$  patterns. Let the subpattern nearest to the probe in the  $i$ 'th subnetwork be  $x_{(i)}^{(l_i)}$ . Denoting  $(y, x) = y^T x$ , we have

$$[Tx]_k = \sum_{i=1}^{S_k} (x_{(i)}^{(l_i)}, x) [x_{(i)}^{(l_i)}]_k + \sum_{i=1}^{S_k} \sum_{\substack{j=1 \\ j \neq l_i}}^{M_i} (x_{(i)}^{(j)}, x) [x_{(i)}^{(j)}]_k \quad (10)$$

which can be written as

$$[Tx]_k = \sum_{i=1}^{S_k} (K - 2d[x_{(i)}^{(l_i)}, x]) [x_{(i)}^{(l_i)}]_k + \sum_{i=1}^{S_k} \sum_{\substack{j=1 \\ j \neq l_i}}^{M_i} (K - 2d[x_{(i)}^{(j)}, x]) [x_{(i)}^{(j)}]_k \quad (11)$$

Assuming that there is no conflict between the  $k$ 'th bits of  $x_{(i)}^{(l_i)}$ ,  $i = 1, \dots, S_k$  (otherwise simultaneous convergence to all the stored subpatterns closest to the probe is not possible), and that  $r[x_{(i)}^{(l_i)}, x] < 1/2$ ,  $i = 1, \dots, S_k$ , it can be seen that the distances  $d[x_{(i)}^{(l_i)}, x]$ ,  $i =$

$1, \dots, S_k$  will decrease (if  $x_k \neq [x_{(i)}^{(l)}]_k$ ) or remain the same (if  $x_k = [x_{(i)}^{(l)}]_k$ ) if and only if the sign of  $[Tx]_k$  is the same as that of  $[x_{(i)}^{(l)}]_k$ , which will be the case if and only if the double sum in (11) does not offset the first sum. If a subnetwork contains replicas or converses of the stored subpatterns  $x_{(i)}^{(l)}$ ,  $i = 1, \dots, S_k$ , the corresponding terms will be included in the first sum of (11), increasing its magnitude with respect to the double sum. It follows that if the network approaches the stored subpatterns when no replicas or converses of these subpatterns are stored, it will also do so when replicas and converses are stored.

It seems sensible that each synaptic contact can only have a certain maximal strength, implying that the synaptic parameter values are restricted to be between some extremal values. Furthermore, storage of an unlimited number of patterns would require that the sensitivity of the synapses to a relative change in the number of stored patterns increase with this number in order to maintain accuracy, which does not seem to make physical sense. When the Hebbian storage rule is used, the diagonal elements  $T_{i,i}$  count the stored patterns and storage may terminate when they exceed a certain value. A maximal value of  $T_{max}$  for the synaptic parameters allows for the storage of  $T_{max}/S_k$  subpatterns, on the average, in each of the  $S_k$  nested subnetworks, sharing the  $k$ 'th neuron. This storage constraint would imply that the maximum capacity of a fully connected network is essentially independent of the network size. As shown in the rest of this section, such restriction on the number of stored subpatterns is also suggested by analytical information retrieval considerations and further justifies the fragmentation of large networks into small subnetworks.

### 3.1 Safe Storage

Suppose that each subnetwork stores, at most, two subpatterns. Denoting the subpatterns stored in the  $i$ 'th subnetwork sharing the  $k$ 'th neuron by  $x_{(i)}^{(1)}$  and  $x_{(i)}^{(2)}$  and assuming that the probe consists of the stored subpatterns  $x_{(i)}^{(1)}$ ,  $i = 1, \dots, S_k$ , we have

$$[Tx]_k = S_k K x_k + \sum_{i=1}^{S_k} (x_{(i)}^{(2)}, x_{(i)}^{(1)}) [x_{(i)}^{(2)}]_k \quad (12)$$

Since, for  $x_{(i)}^{(1)} \neq x_{(i)}^{(2)}$ ,  $|(x_{(i)}^{(2)}, x_{(i)}^{(1)})| < K$ , the sign of  $[Tx]_k$  is the same as that of  $x_k$ . It follows that if the number of subpatterns stored in each of the subnetworks does not exceed 2, they are equilibrium points of the network. For  $K=2$ , there can only be two different subpatterns and, unless they are converse, they would be mutually orthogonal, yielding  $(x_{(i)}^{(2)}, x_{(i)}^{(1)}) = 0$ . Hence,  $[Tx]_k = 2S_k x_k$ , implying that the stored subpatterns are equilibrium

points of the network. For  $K=3$ , storage of 3 subpatterns,  $x_{(i)}^{(j)}$ ,  $j = 1, 2, 3, i = 1, \dots, S_k$ , in each subnetwork yields, assuming, as before, that the corresponding subpatterns of the probe are identical to  $x_{(i)}^{(1)}$ ,  $i = 1, \dots, S_k$ ,

$$[Tx]_k = 3S_k x_k + \sum_{i=1}^{S_k} \{ (x_{(i)}^{(2)}, x_{(i)}^{(1)}) x_{(i)}^{(2)} + (x_{(i)}^{(3)}, x_{(i)}^{(1)}) x_{(i)}^{(3)} \} \quad (13)$$

For  $x_{(i)}^{(2)} \neq x_{(i)}^{(1)}$  and  $x_{(i)}^{(3)} \neq x_{(i)}^{(1)}$ , we have  $| (x_{(i)}^{(2)}, x_{(i)}^{(1)}) | = 1$  and  $| (x_{(i)}^{(3)}, x_{(i)}^{(1)}) | = 1$ , implying that the first term of (13) has a greater magnitude than the second and, consequently, the stored subpatterns are equilibrium points of the network. It follows from similar arguments that in the storage of more than 3 subpatterns in each subnetwork, the second term may exceed the first in magnitude and the stored subpatterns cannot be guaranteed to be equilibrium points. Hence, for  $K = 3$ , the stored subpatterns can be guaranteed to be equilibrium points of the network even in the worst case if and only if their number does not exceed 3.

For  $K \geq 4$  let the subpatterns stored in each of the subnetworks sharing the  $k$ 'th neuron be 3 vectors, the first of which has +1 in all positions, the second has -1 in the  $k$ 'th position and +1 elsewhere and the third has +1 in the  $k$ 'th position and -1 elsewhere (only the components of these subpatterns corresponding to the associated subnetworks are mentioned, the others being 0). It can be readily verified that the  $k$ 'th element of  $Tx$  is  $-S_k(K-4)$ , whose sign is different from that of  $x_k = [\sum_{i=1}^{S_k} x_{(i)}^{(1)}]_k = S_k$  for  $K \geq 4$ . Hence, for  $K \geq 4$ , the subpatterns stored in each subnetwork can be guaranteed to be equilibrium points of the network if and only if their number in each subnetwork does not exceed 2. It might be noted that the same results can be obtained for a fully connected network by taking  $S_k = 1$  and  $K = N$ . The present result for the latter case is somewhat tighter than that presented in [14], which states that the worst-case capacity is 2 for  $N \geq 6$ . In order to restrict the storage to 2 subpatterns per subnetwork, the maximal allowable value of the synaptic parameters must be set to  $2S$ , where  $S$  is the nesting degree of the network.

We have seen that for the stored subpatterns to be equilibrium points of the network even in the worst case, a nested network of subnetwork size 2 can store only two subpatterns in each of the subnetworks. A nested network of subnetwork size 3 can store 3 subpatterns in each subnetwork and nested networks of larger subnetwork sizes can store only 2 subpatterns per subnetwork. This may suggest a certain advantage of networks of subnetwork sizes 2 and 3. However, as shown next, such networks have no error correction capability.

Suppose again that the  $k$ 'th neuron is to be updated and that this neuron is a member of  $S_k$  nested subnetworks. Suppose further that two subpatterns, at most, are stored in

each of these subnetworks and denote the stored subpattern closer to the the probe in the  $i$ 'th subnetwork by  $x_{(i)}^{(1)}$  and the other by  $x_{(i)}^{(2)}$ . We have

$$\begin{aligned} [Tx]_k &= \sum_{i=1}^{S_k} (x_{(i)}^{(1)}, x) [x_{(i)}^{(1)}]_k + \sum_{i=1}^{S_k} (x_{(i)}^{(2)}, x) [x_{(i)}^{(2)}]_k \\ &= \sum_{i=1}^{S_k} K(1 - 2r[x_{(i)}^{(1)}, x]) [x_{(i)}^{(1)}]_k + \sum_{i=1}^{S_k} K(1 - 2r[x_{(i)}^{(2)}, x]) [x_{(i)}^{(2)}]_k \end{aligned} \quad (14)$$

Assuming that there is no conflict between the values of the  $k$ 'th neurons in all  $x_{(i)}^{(1)}$ ,  $i = 1, \dots, S_k$  (otherwise, there cannot be a simultaneous convergence to all the subpatterns closest to the probe) and that  $d[x_{(i)}^{(1)}, x] < K/2$ ,  $i = 1, \dots, S_k$ , it can be seen that in the worst case the distances  $d[x_{(i)}^{(1)}, x]$  will decrease (if  $x_k \neq [x_{(i)}^{(1)}]_k$ ) or remain the same (if  $x_k = [x_{(i)}^{(1)}]_k$ ) if and only if

$$\sum_{i=1}^{S_k} (1 - 2r[x_{(i)}^{(1)}, x]) > \sum_{i=1}^{S_k} |1 - 2r[x_{(i)}^{(2)}, x]| \quad (15)$$

It can be readily verified that the latter condition cannot sensibly hold for networks of subnetwork size 2 or 3 (noting that for a single erroneous bit  $r[x_{(i)}^{(1)}, x] = 1/2$ , for  $K = 2$ ,  $r[x_{(i)}^{(1)}, x] = 1/3$ , for  $K = 3$  and assuming  $r[x_{(i)}^{(2)}, x] > r[x_{(i)}^{(1)}, x]$ ). It follows that nested networks of subnetwork sizes 2 or 3 lack any error correction capability. A sufficient condition for (15) is

$$(1 - 2r[x_{(i)}^{(1)}, x]) > |1 - 2r[x_{(i)}^{(2)}, x]| \quad \text{for } i = 1, \dots, S_k \quad (16)$$

yielding the simultaneous conditions

$$r[x_{(i)}^{(2)}, x] - r[x_{(i)}^{(1)}, x] > 0 \quad \text{for } i = 1, \dots, S_k \quad (17)$$

and

$$r[x_{(i)}^{(2)}, x] + r[x_{(i)}^{(1)}, x] < 1 \quad \text{for } i = 1, \dots, S_k \quad (18)$$

A normal situation would be one in which the two subpatterns stored in a subnetwork are substantially different from each other and the corresponding subpattern of the probe is close to one of them (the probe representing a possibly erroneous version of a stored pattern). Typical values are, say,  $r[x_{(i)}^{(1)}, x] = 0.1$  and  $r[x_{(i)}^{(2)}, x] = 0.6$ . Since  $x_{(i)}^{(2)}$  and  $x$  can be assumed, in essence, to be statistically independent, the normalized distance between them,  $r[x_{(i)}^{(2)}, x]$ , will approach 0.5 as the pattern size increases. Condition (18) is, then, more likely to hold as the subnetwork size increases. The probability of condition (18) holding, given that condition (17) does, constitutes a lower bound on the error correction probability

of the network. While the calculation of this probability is generally a combinatorially hard computation problem, it can be performed for small subnetworks. For sizes 5 and 7 (compatible with the square and hexagonal structures of Figures 1(a) and 1(c)) and for  $r[x_{(i)}^{(1)}, x] \leq 0.2$  (that is, one erroneous bit, at most, out of 5 or 7, respectively), the lower bound on the error correction probability was found to be

$$P_5 \geq 0.9689 \quad P_7 \geq 0.9928$$

It can be seen that, even for small subnetworks, storage of only two subpatterns guarantees a high error correction capability. It should also be noted that conditions (17) and (18) yield convergence to the subpatterns closest to the probe in a single update cycle.

### 3.2 Orthogonal Subpatterns

The safe storage capacity of 2 subpatterns per subnetwork may be increased if the storage mechanism is made discriminatory. Since the only information assumed to be available in the network consists of the state variables and the synaptic parameters, any discrimination rule must be based on their values. It can be seen from (6) that a sufficient condition for the stored subpatterns to be equilibrium points of the network is

$$r[x_{(i)}^{(j)}, x_{(i)}^{(l_i)}] = 1/2 \quad \text{or} \quad (x_{(i)}^{(j)}, x_{(i)}^{(l_i)}) = 0 \quad \text{for} \quad j \neq l_i$$

which means that the subpatterns stored in each of the subnetworks are mutually orthogonal in the Euclidean sense. Storage of only orthogonal subpatterns would also eliminate the redundant storage of nearly repetitive ones. It can be accomplished by modifying the storage rule to be

$$T(M_i + 1) = \begin{cases} T(M_i) + x_{(i)}^{(M_i+1)} x_{(i)}^{(M_i+1)T} & \text{if } T(M_i) x_{(i)}^{(M_i+1)} = 0 \\ T(M_i) & \text{otherwise} \end{cases} \quad (19)$$

After one pattern of size  $N$  is stored, the probability that another random pattern of the same size is orthogonal to the first is

$$P_N = \frac{1}{2^N - 1} \binom{N}{N/2}$$

Employing the Stirling approximation (see, e.g., Hamming [20], p. 165)  $N! \simeq N^N e^{-N} \sqrt{2\pi N}$ , we obtain

$$P_N \simeq \frac{2}{\sqrt{2\pi N}} \quad (20)$$



Hence, the probability that the second pattern is orthogonal to the first is, approximately, inversely proportional to the square root of the pattern size. It follows that the orthogonality requirement is impractical for the storage of network-size patterns in large fully connected networks, as almost every pattern following the first will be rejected. On the other hand, for nested networks consisting of relatively small subnetworks, the orthogonality condition allows for the storage of subpatterns with relatively high probability. For instance, the above approximation gives

$$P_6 = 0.3257, \quad P_{10} = 0.2523, \quad P_{100} = 0.0798$$

The fact that orthogonality, which guarantees stability at the stored subpatterns, can be more sensibly accomodated in small networks than in large ones, further motivates the fragmentatin of large networks into small subnetworks.

The error correction capability of the nested network is examined next, under the assumption that the stored subpatterns are mutually orthogonal. Suppose, as before, that the  $k$ 'th neuron is to be updated and that it is a member of  $S_k$  nested subnetworks, the  $i$ 'th of which stores  $M_i$  subpatterns. Let the subpattern nearest to the probe in the  $i$ 'th subnetwork be  $x_{(i)}^{(l_i)}$ . Then assuming, as before, that there is no conflict between the  $k$ 'th bits of  $x_{(i)}^{(l_i)}$ ,  $i = 1, \dots, S_k$  and that  $r[x_{(i)}^{(l_i)}, x] < 1/2$ ,  $i = 1, \dots, S_k$ , it can be seen from (11) that the distances  $d[x_{(i)}^{(l_i)}, x]$ ,  $i = 1, \dots, S_k$  will decrease (if  $x_k \neq [x_{(i)}^{(l_i)}]_k$ ) or remain the same (if  $x_k = [x_{(i)}^{(l_i)}]_k$ ) even in the worst case if and only if

$$\sum_{i=1}^{S_k} K(1 - 2r[x_{(i)}^{(l_i)}, x]) > \sum_{i=1}^{S_k} \sum_{\substack{j=1 \\ j \neq l_i}}^{M_i} |K(1 - 2r[x_{(i)}^{(j)}, x])|$$

It follows that, since the stored patterns are mutually orthogonal, the maximum possible value of  $|1 - 2r[x_{(i)}^{(j)}, x]|$  is  $2r[x_{(i)}^{(l_i)}, x]$  (it is obtained for either  $r[x_{(i)}^{(j)}, x] = 1/2 - r[x_{(i)}^{(l_i)}, x]$  or  $r[x_{(i)}^{(j)}, x] = 1/2 + r[x_{(i)}^{(l_i)}, x]$ ). It follows that convergence to the subpatterns closest to the probe can be guaranteed if and only if

$$\sum_{i=1}^{S_k} (1 - 2r[x_{(i)}^{(l_i)}, x]) > \sum_{i=1}^{S_k} 2(M_i - 1)r[x_{(i)}^{(l_i)}, x] \quad (21)$$

A sufficient condition for (21) is

$$1 - 2r[x_{(i)}^{(l_i)}, x] > 2(M_i - 1)r[x_{(i)}^{(l_i)}, x] \quad \text{for } i = 1, \dots, S_k$$

yielding

$$r[x_{(i)}^{(l_i)}, x] < \frac{1}{2M_i} \quad \text{for } i = 1, \dots, S_k \quad (22)$$

or

$$M_i < \frac{1}{2r[x_{(i)}^{(l_i)}, x]} \quad \text{for } i = 1, \dots, S_k \quad (23)$$

It follows that the correction of errors of a given proportional size  $r$  is guaranteed if each of the subnetworks stores  $M < 1/2r$  patterns. While for a nested network with disconnected layers ( $S_k = 1$ ) the latter condition is not only sufficient but also necessary, it is only sufficient for nested networks with connected layers ( $S_k \geq 1$ ). This implies that inter-layer connections enhance the potential error correction capability of the network. The number of mutually orthogonal subpatterns that can be stored in a subnetwork is restricted, of course, by the subnetwork size. The error correction capability will reduce as the number of stored patterns is increased, as the diagonal elements of the synaptic parameter matrix will become large with respect to the nondiagonal elements.

### 3.3 Random Subpatterns

When no discrimination is applied to the storage of subpatterns, the latter may be regarded as random vectors, whose components are generated as a sequence of Bernoulli trials (see, e.g., [11]). Writing (5) as

$$[Tx]_k = S_k K x_k + \sum_{i=1}^{S_k} \sum_{\substack{j=1 \\ j \neq l_i}}^{M_i} \sum_{m=1}^K [x_{(i)}^{(j)}]_m [x_{(i)}^{(l_i)}]_m [x_{(i)}^{(j)}]_k \quad (24)$$

where  $[x_{(i)}^{(j)}]_m$  and  $[x_{(i)}^{(l_i)}]_m$ ,  $m = 1, \dots, K$  are the elements of  $x_{(i)}^{(j)}$  and  $x_{(i)}^{(l_i)}$ , respectively, corresponding to the  $i$ 'th subnetwork, it can be seen that the triple sum on the right hand side of (24) is composed of  $K \sum_{i=1}^{S_k} (M_i - 1)$  independent random variables, taking on the values  $\pm 1$ . Suppose for simplicity that  $S_k = S$  and  $M_i = M$ ,  $i = 1, \dots, S$ . Then for large  $K$  and  $M$  the second term on the right-hand side of (24) is asymptotically normal with mean zero and variance  $SK(M - 1)$ . It follows that the probability that the second term will not exceed the first is, approximately for large  $K$  and  $M$ , given by

$$P = \frac{1}{2} + \text{erf} \left\{ \left( \frac{SK}{M - 1} \right)^{1/2} \right\} \quad \text{where} \quad \text{erf}\{y\} = \int_0^y e^{-z^2/2} dz$$

Since  $\text{erf}\{y\}$  is a monotone-increasing function, the probability that  $x_k$  will not change increases with  $S$ . It follows that the probability that the stored subpatterns are equilibrium points of the network increases with the nesting degree.

Next, suppose that the probe does not necessarily consist of stored subpatterns, and that the subpattern nearest to the probe in the  $i$ 'th subnetwork is  $x_{(i)}^{(l_i)}$ . Writing (11) as

$$[Tx]_k = \sum_{i=1}^S (K - 2d[x_{(i)}^{(i)}, x])[x_{(i)}^{(i)}]_k + \sum_{i=1}^S \sum_{\substack{j=1 \\ j \neq i}}^{M_i} \sum_{m=1}^K [x_{(i)}^{(j)}]_m [x_{(i)}]_m [x_{(i)}^{(j)}]_k \quad (25)$$

where  $[x_{(i)}]_m, m = 1 \dots, K$  are the elements of  $x$  corresponding to the  $i$ 'th subnetwork, it follows from arguments similar to the above that the probability that the distances  $d[x_{(i)}^{(i)}, x]$  will decrease (if  $x_k \neq [x_{(i)}^{(i)}]_k$ ) or remain the same (if  $x_k = [x_{(i)}^{(i)}]_k$ ) is approximately, for large  $K$  and  $M$ ,

$$P = \frac{1}{2} + \operatorname{erf} \left\{ \left( \frac{K \sum_{i=1}^S (1 - 2r[x_{(i)}^{(i)}, x])}{M - 1} \right)^{1/2} \right\}$$

which implies that the probability of error correction generally increases with the nesting degree.

### 3.4 On the Retrieval of Patterns and Subpatterns

The nested network, probed by any pattern, is expected to converge to the stored subpatterns nearest to the corresponding subpatterns of the probe. These subpatterns may each constitute meaningful information or may jointly form larger subpatterns or network-size patterns. Even if very few subpatterns are stored in each subnetwork, these subpatterns can make up a vast number of larger patterns. To illustrate this point, suppose that the nested network has a total number of  $N$  neurons and a subnetwork size  $K$ , hence, at least  $N/K$  subnetworks ( $N/K$  being the number of subnetworks in the lowest layer). If the storage capacity of each subnetwork is  $C$ , the total capacity of the network is  $C^{N/K}$  network-size patterns. For  $N = 10^3$ ,  $K = 10$  and  $C = 2$ , this capacity is  $2^{100}$ , which is vastly greater than, not only the safe capacity of 2 for fully connected networks, but also any known upper bound on the capacity of such networks. If patterns stored in subnetworks or in larger sections of the network can be read individually, the storage capacity is vastly greater yet.

How can subpatterns be read or corrected? Suppose that each subnetwork stores the "null" subpattern, consisting of  $-1$ 's in all the positions corresponding to the subnetwork and  $0$ 's elsewhere, in the initial values of the synaptic parameters and that any additional subpattern will only be stored if it is not null. Then when a subnetwork is probed by a null subpattern, it will also produce it as an output. The subnetwork can then be said to be inactivated. Reading an individual subnetwork or a group of subnetworks means that all the other subnetworks are inactivated. The probe may be viewed as a network-size pattern,

in which many or most of the subpatterns are null. It activates a certain set of subnetworks, corresponding to the subpatterns which are not null. Each of the affected subnetworks, in turn, converges to the stored subpattern nearest to the corresponding subpattern of the probe and the network as a whole produces the stored information corresponding to the probe. The network is then a memory device, in which there is a close correspondence between information and the address in which it is stored.

## 4 Examples

The nested network depicted in Figure 2, which consists of 85 neurons, has been used for illustration purposes. A fully connected network of the Hopfield type was also defined for comparison purposes, by connecting each of the neurons to all the other neurons, using the same sampling (by the same neurons) of the pattern field. (Both the cases of self-connected and self-disconnected neurons in the latter structure were examined; all of the results obtained were identical for the two cases).

Four patterns were stored in the two networks. Because of the small size of the subnetworks, storage in the nested network was restricted to two subpatterns per subnetwork, including the null subpattern, by restricting the maximal synaptic parameter values to 4 (as the nesting degree is 2). The networks were then probed by each of the patterns and the final patterns which the networks settled into were recorded. The results are shown in Figure 3. Part (a) of the figure shows the neurons that were activated by the probe (receiving the value +1), represented by dots. Parts (b) and (c) show the final states of the fully connected and the nested networks, respectively. It can be seen that while the fully connected network has "wiped out" the patterns, they were perfectly maintained by the nested network. The patterns were then corrupted by inverting the sign of every 5<sup>th</sup> neuron and the two networks were probed by the resulting patterns. The probes are shown in part (d) of Figure 3 and the final states of the fully connected network and the nested one are shown in parts (e) and (f), respectively. It can be seen that, in spite of the high error rate, the original patterns were essentially recovered by the nested network. This, again, stands in contrast to the performance of the fully connected network, which eliminates the information.

It might be suggested that, for the fully connected network to perform well, the patterns should occupy a larger portion of the pattern field (see, e.g., Lippman [21]). To examine this possibility and, at the same time, illustrate the reconstruction of large

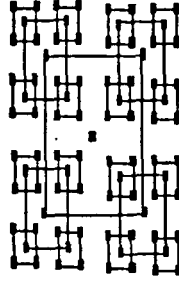


Figure 2: The nested network used in the examples

patterns from small stored subpatterns, the patterns A-J, depicted in Figure 4, were stored in the two networks. For the nested network, storage was restricted, as before, to 2 subpatterns per subnetwork, including the null subpattern. Each of the networks was then probed by each of the letters and the final patterns which the networks settled into were recorded. The average number of errors for the 10 patterns was 25.2 for the fully connected network and 8.1 for the nested one. The patterns are shown in Figure 5 for the letters A and J. It can be seen that, even for this small number of stored patterns, which conforms with the known capacity bounds for the Hopfield network ([8], [11]), the fully connected network shows a considerably lower level of stability at the stored patterns than the nested one. The performance of the nested network in this case is particularly remarkable in view of the fact that it is required to reconstruct full network-size patterns (say, J) from stored subpatterns of other patterns (say, A, B), as only the first subpattern presented to each subnetwork, following the null subpattern, is stored.

## 5 Conclusion

Nested neural networks allow for the storage and retrieval of patterns of different sizes, which may individually constitute meaningful information or jointly form larger patterns. Restricted storage of only few subpatterns in each subnetwork results in a vast storage capacity for the entire network, maintaining high degrees of stability and error correction capability. This paper has considered only the possible role of a nested neural network as a memory device for predetermined patterns of neural states. The input-output interface of sensory information with such networks is the subject of present research.

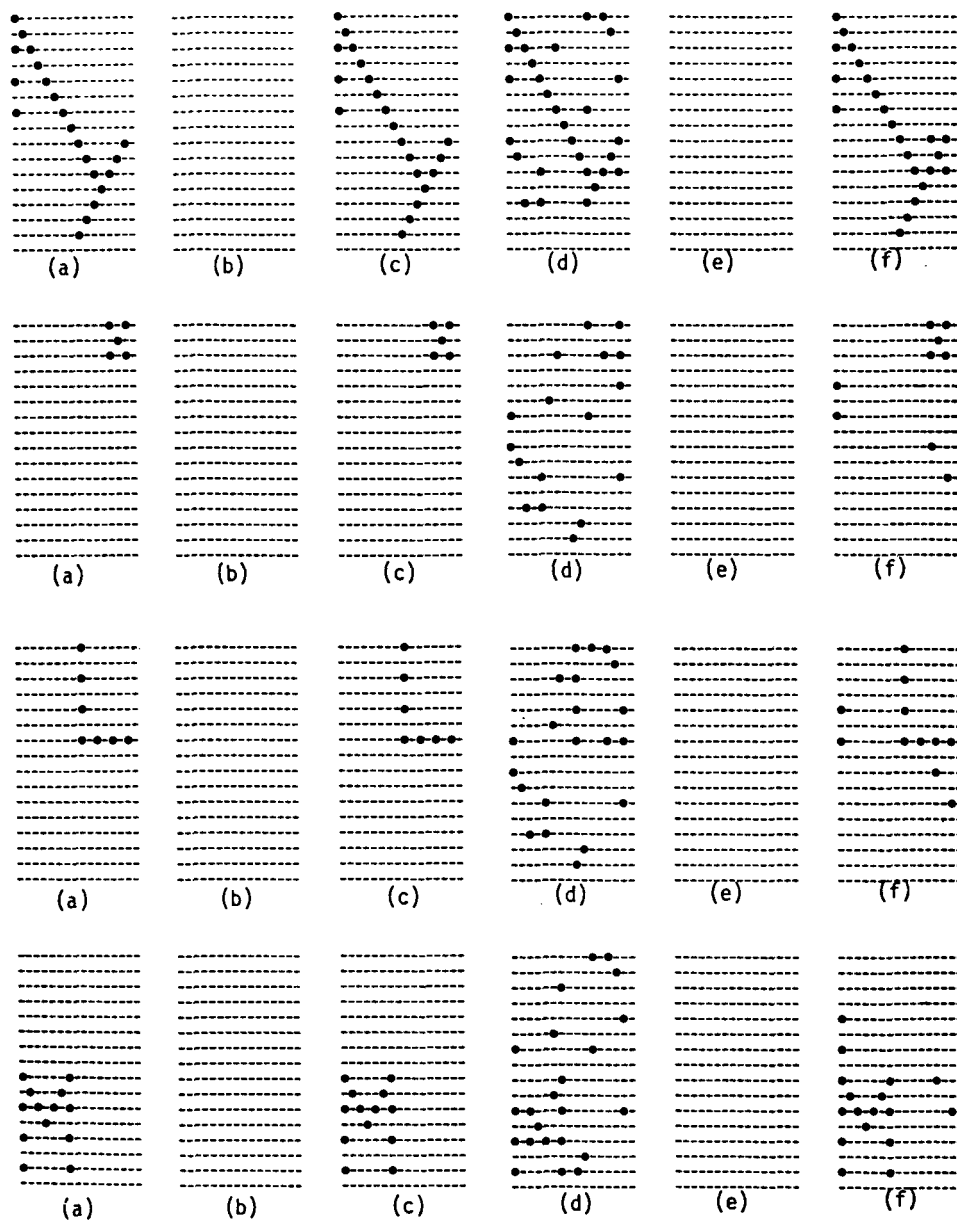


Figure 3: The final states of the fully connected network (b), (e) and of the nested network (c), (f) in response to a probe (a) and a corrupted version of it (d)).

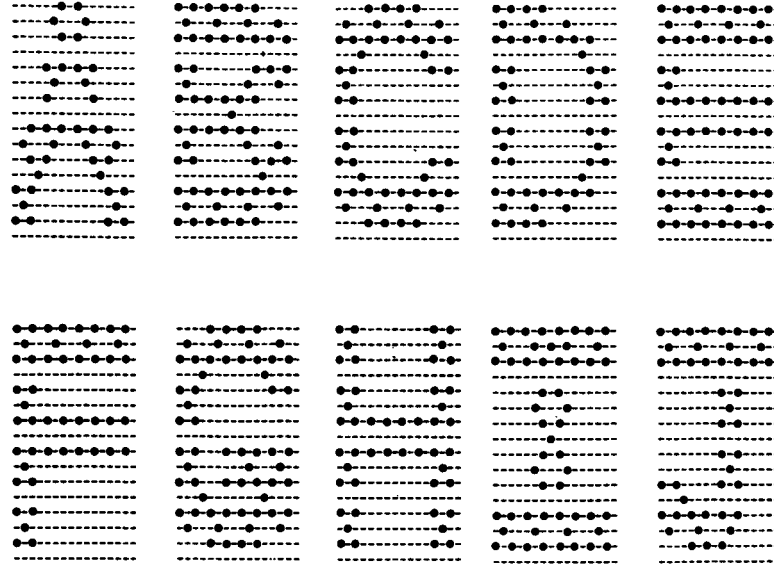


Figure 4: The patterns A-J stored in the two networks.

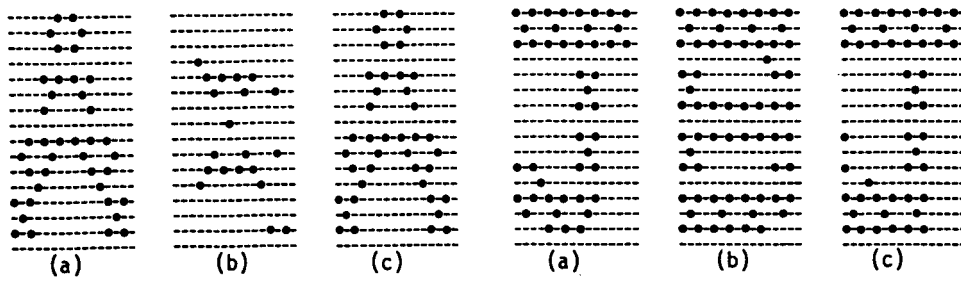


Figure 5: The final patterns of the fully connected network (b) and of the nested network (c), storing the patterns A-J and probed by the patterns A and J (a).

## References

- [1] F. Crick and C. Asanuma, "Certain Aspects of the Anatomy and Physiology of the Cerebral Cortex," in J. L. McClelland, D. E. Rumelhart and the PDP Research Group, *Parallel Distributed Processing*, Vol. 2, The MIT Press, Cambridge, MA, 1987.
- [2] F. Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Spartan Books, Washington, DC, 1961.
- [3] T. Kohonen, *Associative Memory – A System Theoretic Approach*, Springer Verlag, New York, 1977.
- [4] S. Grossberg, *Studies of Mind and Brain*, D. Reidel, Boston, 1982.
- [5] D.O. Hebb, *The Organization of Behaviour*, Wiley, New York, 1949.
- [6] L. N. Cooper, "A Possible Organization of Animal Memory and Learning," in *Proc. Nobel Symposium on Collective Properties of Physical Systems*, Vol. 24, B. Lundquist and S. Lundquist, eds., pp. 242-264, Academic Press, New York, 1973.
- [7] W.S. McCulloch and W. Pitts, "A Logical Calculus of the Ideas Imminent in Nervous Activity," *Bull. Math. Biophys.*, Vol. 5, pp 115-133, 1943.
- [8] J.J. Hopfield, "Neural Networks and Physical Systems with Emergent Computational Abilities," *Proc. Nat. Acad. Sci. USA*, Vol. 79, pp. 2554-2558, April 1982.
- [9] J.J. Hopfield, "Neurons with Graded Response have Collective Computational Properties like those of Two-State Neurons," *Proc. Nat. Acad. Sci. USA*, Vol. 81, pp. 3088-3092, May 1984.
- [10] Y. S. Abu-Mostafa and J-M St.Jacques, "Information Capacity of the Hopfield Memory," *IEEE Trans. on Information Theory*, Vol. IT-31, No. 4, July 1985.
- [11] R.J. McEliece, E.C. Posner, E.R. Rodemich and S.S. Venkatesh, "The Capacity of the Hopfield Associative Memory," *IEEE Trans. on Information Theory*, Vol. IT-33, No. 4, July 1987.
- [12] A.D. Bruce, E.J. Gardner and D.J. Wallace, "Dynamics and Statistical Mechanics of the Hopfield Model," *J. Phys A: Math. Gen.*, Vol. 20, pp. 2909-2934, 1987.



- [13] B. L. Montgomery and B. V. K. Vijayakumar, "Evaluation of the use of the Hopfield Neural Network Model as a Nearest Neighbor Algorithm," *Appl. Opt.*, Vol. 25, pp. 3759-3766, 1986.
- [14] A. Dembo, "On the Capacity of Associative Memories with Linear Threshold Functions," submitted for publication.
- [15] D. Horn and J. Weyers, "Information Packing in Associative Memory Models," *Physical Review A*, Vol. 34, No. 3, pp. 2324-2328, September 1986.
- [16] B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Company, New York, 1983.
- [17] D. E. Dudgeon and R.M. Mersereau, *Multidimensional Digital Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1984.
- [18] A. J. Ahumada, Jr., "Simulated Annealing in Networks for Computing Possible Arrangements for Red and Green Cones," *Proc. of the IEEE First Internat. conf. on Neural Networks*, Vol. IV, pp. 107-114, San Diego, CA, June 1987.
- [19] A. B. Watson and A.J. Ahumada, Jr., "An Orthogonal Oriented Quadrature Hexagonal Image Pyramid," *NASA Technical Memorandum 100054*, December 1987.
- [20] R. W. Hamming, *Coding and Information Theory*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1986.
- [21] R. P. Lippman, "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, pp. 4-22, April 1987.



## Report Documentation Page

1. Report No. NASA TM-101032		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle  Nested Neural Networks				5. Report Date August 1988	
				6. Performing Organization Code	
7. Author(s)  Yoram Baram				8. Performing Organization Report No. A-88288	
				10. Work Unit No. 505-66-11	
9. Performing Organization Name and Address  Ames Research Center Moffett Field, CA 94035				11. Contract or Grant No.	
				13. Type of Report and Period Covered Technical Memorandum	
12. Sponsoring Agency Name and Address  National Aeronautics and Space Administration Washington, DC 20546-0001				14. Sponsoring Agency Code	
15. Supplementary Notes  Point of Contact: Dallas G. Denery, Ames Research Center, MS 210-9 Moffett Field, CA 94035 (415) 694-5427 or FTS 464-5427					
16. Abstract  Nested neural networks, consisting of small interconnected subnetworks, allow for the storage and retrieval of neural state patterns of different sizes. The subnetworks are naturally categorized by layers corresponding to spatial frequencies in the pattern field. The storage capacity and the error correction capability of the subnetworks generally increase with the degree of connectivity between layers (the "nesting degree"). Storage of only few subpatterns in each subnetwork results in a vast storage capacity of patterns and subpatterns in the nested network, maintaining high stability and error correction capability.					
17. Key Words (Suggested by Author(s)) Neural networks Fractal networks Pattern processing Associative memory			18. Distribution Statement Unclassified-Unlimited  Subject Category - 63		
19. Security Classif. (of this report)  Unclassified	20. Security Classif. (of this page)  Unclassified		21. No. of pages  25	22. Price  A02	